ShadowPuppets: Supporting Collocated Interaction with Mobile Projector Phones Using Hand Shadows

Lisa G. Cowan
Computer Science and Engineering
University of California, San Diego
lgcowan@cs.ucsd.edu

Kevin A. Li
AT&T Labs – Research
Florham Park, NJ
kevinli@research.att.com

ABSTRACT

Pico projectors attached to mobile phones allow users to view phone content using a large display. However, to provide input to projector phones, users have to look at the device, diverting their attention from the projected image. Additionally, other collocated users have no way of interacting with the device.

We present ShadowPuppets, a system that supports collocated interaction with mobile projector phones. ShadowPuppets allows users to cast hand shadows as input to mobile projector phones. Most people understand how to cast hand shadows, which provide an easy input modality. Additionally, they implicitly support collocated usage, as nearby users can cast shadows as input and one user can see and understand another user's hand shadows.

We describe the results of three user studies. The first study examines what hand shadows users expect will cause various effects. The second study looks at how users perceive hand shadows, examining what effects they think various hand shadows will cause. Finally, we present qualitative results from a study with our functional prototype and discuss design implications for systems using shadows as input. Our findings suggest that shadow input can provide a natural and intuitive way of interacting with projected interfaces and can support collocated collaboration.

Author Keywords
Projector-camera system, mobile projector phone, shadow, gesture, interaction technique

ACM Classification Keywords
H5.2 [Information interfaces and presentation]: User Interfaces; B 4.2 Input Output devices

General Terms
Design, Experimentation, Human Factors

INTRODUCTION

Sharing information displayed on a mobile device's small screen with collocated people can be difficult. Pico projectors make it easier for mobile phone users to share visual information with those around them using a projected image, which can be much larger than the device's screen. However, current commodity projector phones only support input via the handset's user interface. As a result, users must look at the handset to interact with the phone's buttons or touch screen, dividing attention between the handset and the projected display. This context switching can distract presenters and viewers from ongoing conversations taking place around the projected display. Additionally, viewers may find it difficult to interpret what the presenter is doing as he interacts with the handset, and they have no way of interacting with the system themselves.

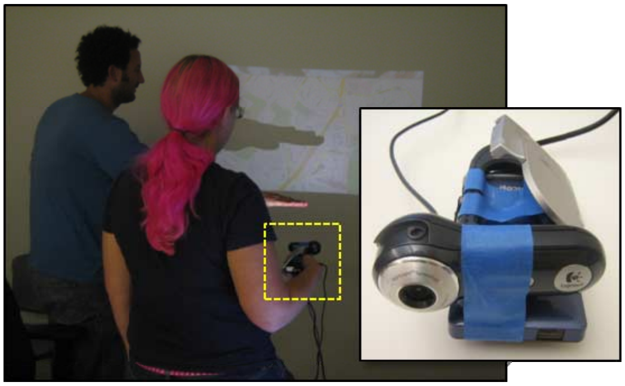

Figure 1: Interacting with ShadowPuppets prototype.

We present ShadowPuppets, a mobile projector phone system that allows users to provide input to the system by casting shadows (Figure 1). Users hold the mobile projector phone in one hand while casting shadows with the other hand, and shadows are detected using an attached camera. ShadowPuppets, like some previous research systems [3,4], precludes the need for visual attention to the device and supports ad hoc interaction with uninstrumented environments. Additionally, shadow interaction enables bystanders to provide input by casting hand shadows, without requiring additional equipment. Potentially, multiple collocated users could interact simultaneously.

We conducted a formative survey asking 19 smartphone users about the mobile phone applications they commonly use when collaborating with others in collocated settings. The maps and photo browsing applications were the most popular applications in these settings. These results formed the basis for guiding the design of shadows for our ShadowPuppets prototype.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
Copyright 2011 ACM 978-1-4503-0267-8/11/05....$10.00.

Our first user study was focused on determining what types of shadows such a system should support, from the perspective of a user trying to interact with the system. In this study, we showed our participants video clips of different effects for both Maps and PhotoBrowser applications that would happen in response to a hypothetical user action (i.e. panning left on a map), though no shadow was shown. Participants were then asked to produce the shadow that they felt most appropriate for causing that effect. The shadows we observed provided an initial set of shadows to explore.

Since ShadowPuppets was motivated as a method for collocated collaboration, we were also interested in how observers would interpret shadows. Our second user study focused on this observer perspective. We presented participants with videos of hand shadows and then asked them to determine what kinds of effects the actor in the video was trying to cause.

In a final qualitative study, we documented how collocated users interacted with our functional ShadowPuppets prototype and reflected on their experiences. Our findings suggest that shadows can provide a natural and intuitive way of interacting with projected interfaces and can support collocated collaboration.

RELATED WORK

We build on related work on mobile projector camera systems and gestural computing. We also consider how shadows have been leveraged in interactive systems, and discuss relevant research on awareness in groupware.

Augmenting the Environment with Mobile Projectors

The seminal *Everywhere Displays* system [22] used steerable projectors to create multiple interactive surfaces within an environment. Researchers later investigated techniques for interacting with handheld projected displays: moving and pressing buttons on the device [1], touching the projection surface with fingers [29] or moving the projector like a "*spotlight*" within a virtual information space [23]. Cao et al. expanded on this metaphor, employing passively tracked pens and projectors to define and interact with information spaces and creating techniques for collocated collaboration with multiple projectors [3,4].

Researchers have also explored projector-device ensembles as a way of increasing interactive space. *Bonfire* [17] integrates projectors with a laptop to extend the interactive space to the table. Similarly, *PenLight* [27] and *MouseLight* [28] increase the interactive space for digital pens, providing visual feedback for interaction with paper.

Harrison et al. [12] analyzed vibrations to detect taps on the skin, and demonstrated using their technique to interact with an interface projected on the body, and Mistry et al. [19] relied on computer vision for interaction with a wearable camera-projector system. In the projector-phone space, Greaves and Rukzio [10] found that users preferred projector-based interaction over phone-based interaction. Cowan et al. documented commodity projector phone use "in the wild" [6], finding that, even without projection-specific input techniques, these devices afforded novel interaction modalities.

Gestural Interaction Techniques

Gestural interaction addresses many of the issues we examine. A mobile touchscreen, such as the *Touch Projector,* can be used to interact with distant displays but requires all users to have a handheld device [2]. Computer vision has been used for un-instrumented detection of gestures, such as pointing [5] and pinching [32]. Gustafson et al. [11] explore screenless, spatial, gestural interaction with *Imaginary Interfaces*, relying on users' memory in lieu of visual feedback. However, it can be difficult for observers to interpret a user's intent based on his gestures, without visual feedback, and interpretability is important for collaboration.

Usage of shadows in interactive computer systems

Shadows have been employed in interactive systems for both input and output because they are intuitively understood by users. Shoemaker et al. explored using real and virtual shadow representations of users to extend user's reach and enhance the interpretability of indirect interaction with large wall displays [25], and Snibbe and Raffle [26] analyzed the use of virtual shadow silhouettes to represent participants in interactive museum exhibits. Hilliges, et al. motivate the use of shadow feedback for 3D interaction above tabletops [13] and *ShadowGuides* [9] employed virtual shadows, visual representations of user's raw input, to guide users in learning tabletop gestures.

Computer vision techniques have been developed for robustly detecting shadows [14] and shadow detection has been leveraged for estimating 3D hand positions [24], detecting surface contacts [16], and tracking above-the-surface interactions with multi-touch tables [7].

Awareness and presence in groupware

Much work has focused on awareness and presence in groupware, to support anticipation and interpretation of others' intentions and actions [1]. The *VideoArms* system [30], which employed representations of users' arms, emphasized the role of embodiment for awareness in collocated and remote collaboration and Pinelle et al. [21] found that relative interaction and an arm-like virtual embodiment were preferred for tabletop groupware. Ishii et al. [15] emphasized the value of "*gaze awareness*," awareness of what one's partner is looking at and attending to. Also, several systems, including *VideoWhiteboard* [31] and *Distributed Designers' Outpost* [8], have used shadows superimposed on the work surface to provide awareness of remote collaborators' activities.

WHAT ACTIVITIES DO PEOPLE DO COLLABORATIVELY ON THEIR MOBILE PHONES?

We interviewed 19 volunteers (9 female), aged 21–56 (median 31.5) from within our institution about their smartphone usage habits. This survey was intended to guide the design of the ShadowPuppets system by highlighting potential usage scenarios. Interviews lasted 15 minutes. During the interviews, we asked users how often they perform various activities on their smartphones. We also asked how often they perform these activities collaboratively, in collocated settings. Finally, we asked participants to pick the top three activities that were most important to them in the context of collocated collaboration, and asked them to describe some recent scenarios.

Figure 2 plots the activities that were rated most important by respondents for collocated collaboration. Viewing photos, using maps, taking photos, and searching the web rated highest.

Since photo and map viewing were rated as the most important and most frequently performed collocated collaborative activities, we decided to study those applications further in the context of ShadowPuppets. It is less readily apparent how photo capture and web search (which requires extensive text entry) could benefit from gestural interaction with a projected display, even though these were highly rated as well.

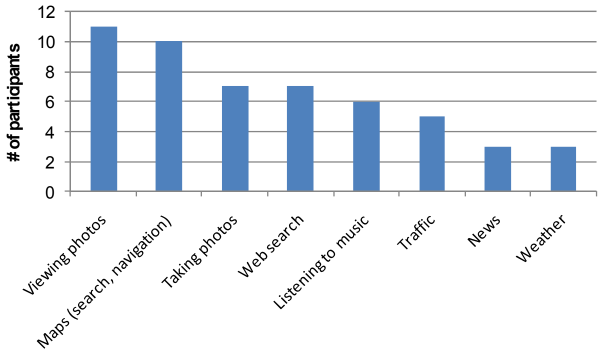

Figure 2. Number of participants out of 19 who rated the respective feature as "most important" for collocated collaboration on their mobile phone.

Current strategies for doing things together on phones

Respondents reported strategies for collocated collaboration on mobile phones varied widely. Their strategies fell into four categories: (1) one person controls the phone and verbalizes information for others, (2) one person holds up the phone for others to view, (3) the phone is passed around, and (4) someone shares a link or reference and others view the information their own phones. Strategies were selected based on group size, closeness of participants, availability of technology, the nature of the information, and other contextual factors. Respondents also varied in their perceptions of device control and privacy: some would physically pass around their handsets while others guarded their mobile phones closely.

USER STUDY 1: WHAT KIND OF HAND SHADOWS SHOULD BE USED?

One of the motivations for using hand shadows as a gestural input into our system was that users already know how to generate shadows with their hands. Although all users know *how* to cast different shadows, it was unclear *what* shadows users would expect to cause the different effects that an application might support.

The purpose of this study was to see what kinds of shadows users would expect to cause various effects in the interface. We structured our study based on Wobbrock et al.'s study of user generated gestures for surface computing [33]. In this study, participants were shown video clips of effects (e.g., a map panning left) for a *Maps* and a *PhotoBrowser* application and asked to generate hand shadows they felt would cause those effects.

Interfaces and Operations

Based on our survey results, we chose *Maps* and *PhotoBrowser* as our interface conditions. For each of these conditions, we considered 7 operations: 4 *Pan* operations (*Up, Down, Left, Right*), 2 *Zoom* operations (*In, Out*), and *Selection*. We chose these operations because they are commonly performed by users interacting with these applications, and because they could be readily conceptually mapped to spatial gestures.

For the *PhotoBrowser* application, *Pan Left* and *Pan Right* are used to view the previous and next photo, *Pan Up* and *Pan Down* are used to scroll through a list of thumbnails, *Zoom In* and *Zoom Out* are used to change the view of a photo, and *Select* may be used to choose a thumbnail or interact with a menu. Similarly, in the *Maps* application, users can *Pan* (we constrain the operation to 4 directions), *Zoom In* or *Out* on the map, and *Select* a marker or icon on the map or a menu option.

While shadow gestures could potentially support more numerous operations, we constrained the study to existing mobile applications, rather than exploring the breadth of possible interactions. Future work could unlock the potential of the ShadowPuppets approach. A primary challenge for scaling this technique to a larger number of gestures is to design gestures that the vision system can recognize and that users can readily learn, perform, and interpret.

Task

Participants were shown a video of an *effect* of what might happen in response to some user action. They were then prompted to cast the shadow that they felt would be most appropriate for causing that effect.

Immediately after casting the shadow, participants were asked to rate two statements on a 7-point Likert scale (1=disagree, 7=agree). The first read, *"The shadow I cast was a good match for causing this effect."* The second read, *"The shadow I cast was easy to make."*

Apparatus

A C# program was used to present videos of the effects of gestures to users. We used a *Microvision SHOWWX* laser pico projector to display the videos. The users and their shadows were video recorded.

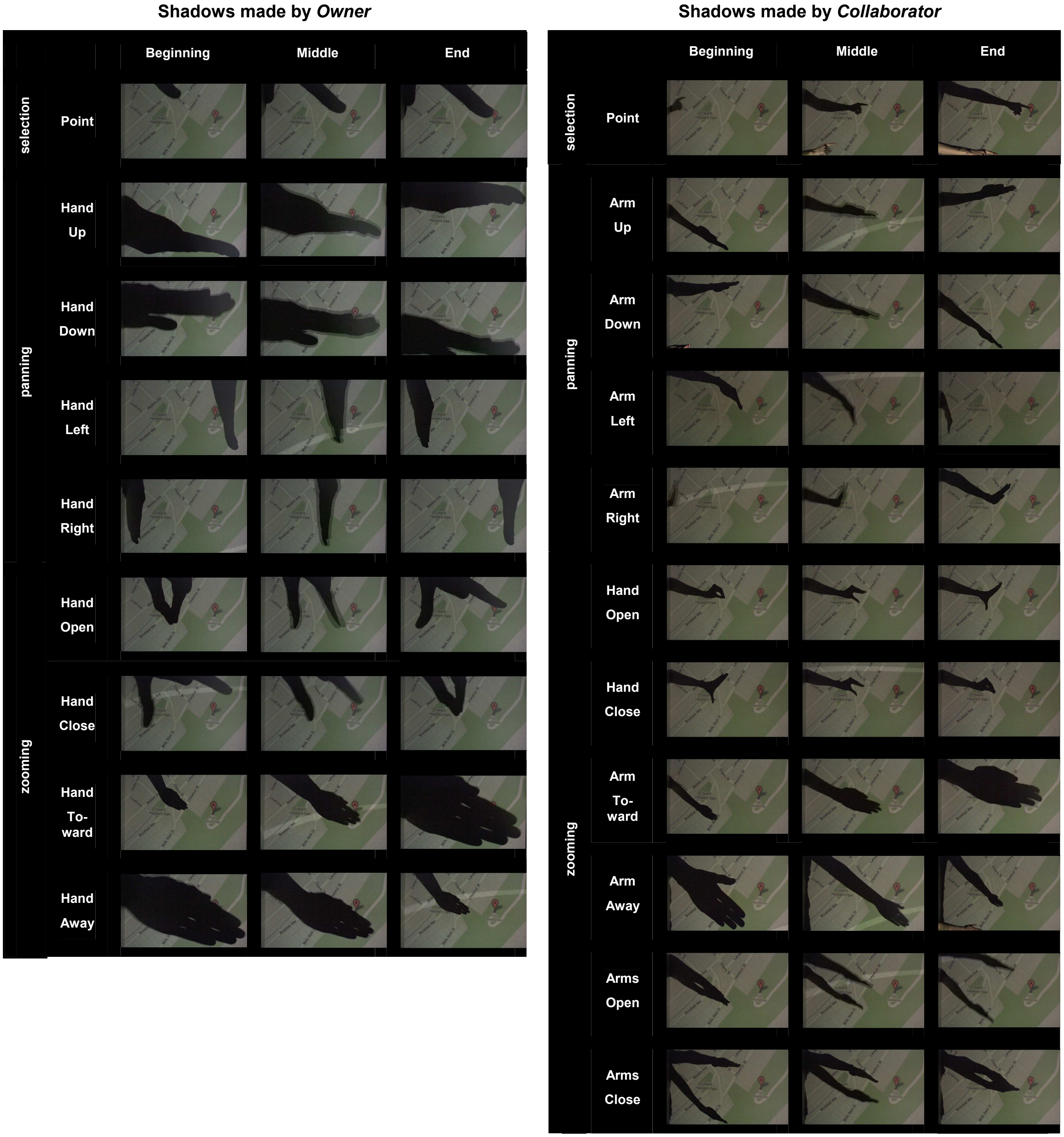

Table 1. The most common shadows generated by users for panning, zooming and selection. (Left) shows the shadows made in the *Owner* condition. (Right) shows the shadows made in the *Collaborator* condition. These are representative of shadows generated by users in Study 1 and formed the basis for the video clips used in Study 2.

Participants

Sixteen volunteers (9 female) ranging in age from 24 to 56 years (median 28.5) were recruited from within our institution. All participants were right-handed. We did not recruit left-handed participants because we expected that some of the tasks might be harder, depending on which hand was holding the projector. Half of our participants owned mobile phones with multitouch capabilities.

Hypothesis

We expected that most participants would cast similar shadows for any given effect. We also hypothesized that owners of multitouch phones would cast different types of shadows, given their experience with gestural interfaces.

Experimental Design

We used a within-participants factorial design. The independent variables were *Applications* (*Map* and *Photo-*

Browser), *Roles* (*Owner* and *Collaborator*), and *Effects* (*Pan Up, Pan Left, Pan Right, Pan Down, Zoom In, Zoom Out, and Select*).

Presentation of *Application* was counterbalanced across participants. We considered mixing tasks from the two *Application* conditions together. However, that might have caused users to converge on the same set of gestures for both contexts, a behavior that we were independently interested in.

For each application, we considered two *Role* conditions: *Owner* and *Collaborator*, with order of presentation counterbalanced across participants. In the *Owner* condition participants held the projector in their left hands and gestured with their right hands, standing 6 feet from the projection surface. In the *Collaborator* condition participants did not hold the projector and stood forward and to the left of the projector, 3 feet from the projection surface. Because C*ollaborators* had both hands free, they could use either or both hands to cast shadows. Within each condition, the various *Effects* were presented in random order.

In summary, the experimental design was: 2 *Applications* (*Map* and *PhotoBrowser*) × 2 *Roles* (*Owner* and *Collaborator*) × 7 *Effects* (*Pan Up, Pan Down, Pan Left, Pan Right, Zoom In, Zoom Out* and *Select*). This resulted in 28 videos shown to 16 participants for a total of 448 shadows.

Goodness of Match and Ease of Making Gesture

We examined the participant ratings using the Mann-Whitney test. Participants rated their shadows in the *Collaborator* condition (*median* = 6) with significantly higher goodness ratings (U = 21468.00, z = 2.74, p < .01) than in the *Owner* condition (*median* = 6). They also rated their shadows in the *Collaborator* condition (*median* = 7) with significantly higher easiness ratings (U = 21201.50, z = 2.944, p < .01) than in the *Owner* condition (*median* = 6).

Owners of multitouch mobile phones rated their shadows significantly higher on the goodness rating (*median* = 6, U = 21201.50, z = 2.944, p < .01) than non-owners (*median* = 6). Owners of multitouch devices also rated their shadows significantly easier (U = 20102.00, z = 3.821, p <.001) than non-owners (*median* = 6).

Most Common Gestures

Here we highlight some of the more common shadows made in response to the different effects. The most popular gestures are highlighted in Table 1. In general, participants would perform similar gestures for a particular effect regardless of the role of *Owner* or *Collaborator*.

Collaborator

Pan: Participants would move their hand or arm across the projection in the direction they wanted to pan. Twelve of 16 participants performed this type of gesture. This is shown as *Arm Up/Down/Left/Right* in Table 1.

Zoom: Participants were split on their zooming technique when standing near the wall. Nine of 16 participants performed large opening and closing gestures with their arms (*Arms Open/Close*). Three participants performed an opening or closing of their hand (hand open/close). Four participants would move their hand toward or away from the wall (*Arm Toward/Away*).

Select: Fourteen participants performed some variant of pointing for selection (*Point*). Of these fourteen, ten would hold their finger near the marker and dwell, while others would wiggle or tap their finger to indicate selection.

Owner

Pan: Similar to the *Collaborator* condition, 12 participants moved their hand in the direction they wanted to pan (*Hand Up/Down/Left/Right*).

Zoom: Eleven participants performed some form of pinching (*Hand Open/Close*) to zoom in and out. Four participants moved their hand toward or away from the wall (*Hand Toward/Away*) to zoom.

Select: Participants used the same form of selection as in the *Collaborator* condition, with 14 participants performing some form of pointing (*Point*).

Discussion

The results of this study show that there was high agreement for many gestures, and similar shadows were cast for both the *Maps* and *PhotoBrowser* applications. Yet multiple gesture aliases may be helpful for supporting different usage scenarios (e.g. right or left-handed) when there is low agreement [33]. For example, there was some disagreement among users regarding pan direction, e.g. some users wanted to move their hands up to *Pan Up* while others wanted to move their hands down to cause the same action. Also, some users thought of zooming as expanding / contracting (performed with pinching gestures) while others thought of it as pushing / pulling (performed by moving hands or arms toward or away from the projector). In cases like this, the designer can either make an arbitrary choice or provide aliases from which users can choose.

Users typically wanted to perform symmetric gestures to perform inverse operations, but this was not always comfortable. Moving the right hand from left to right to *Pan Right* (*Owner* condition) was difficult when holding the projector in the left hand. Similarly, trying to move the arm right to left can be difficult when standing on the left side of the projection (*Collaborator* condition).

Most participants preferred the *Collaborator* condition (14 for *Maps* and 13 for *Photos*). They preferred *Collaborator* because their hand shadows were smaller relative to the display. As a result, they were able to control their shadows with finer granularity, and their shadows occluded less of the display. This was also reflected in significantly higher goodness and easiness ratings. Participants also described *Collaborator* as more “intuitive” since their shadows typically remained within the frame. In contrast, when they were holding the projector in the *Owner* condition, they had to first locate the projection frame and then position their hand appropriately. Participants also described *Col-*

laborator as more "familiar" and "comfortable." The participants who preferred *Owner* cited feeling more in control.

Even though we told participants that we were studying how to use shadow gestures for interaction, many of them described the experience as "*like using a touch screen*". They viewed shadows as a kind of virtual touch on the projection surface, i.e., the shadows were the contact points on a virtual, distant touchscreen.

Users typically mentioned familiar metaphors to describe their shadow gestures, which seemeed to provide some benefits of virtual reality. They drew on the direct manipulation GUI paradigm ("*using a trackball or mouse*", "*clicking*", "*using a scrollbar*") and on the physical world ("*grasping*," "*pushing*," "*pulling*," "*pinching*," "*tapping*," "*clicking*," and "*pointing*"). Only one user made a symbolic gesture (drawing a "z" with her finger to zoom).

In many cases, our participatory design approach resulted in direct translations of familiar multi-touch gestures, drawn from participants' prior knowledge and experiences. Direct translations, however, may sometimes be appropriate and are common starting points in new domains.

USER STUDY 2: DO USERS UNDERSTAND OUR LANGUAGE OF HAND SHADOWS

The previous study helped guide the set of shadows that our system should support. However, it remained to be seen whether casual observers would be able to determine the desired effect when shown a shadow gesture. As we alluded to earlier, this would be a potential benefit of using shadows for collocated interaction.

The purpose of this study was to see whether participants would be able to identify what action a user was trying to perform based on his hand shadow.

We also hoped to gain insight on an observation from the previous study regarding different shadows. For certain effects, multiple types of shadows were observed. Were some shadows more understandable than others?

In the previous user study we showed users an effect and elicited the gesture that would cause it. In this study, we did the inverse. We presented participants with a video clip of a shadow being cast and asked them what effect they thought it would cause. We were motivated to run this follow-up study, as suggested by Wobbrock, et al. [33], to gauge the intuitiveness of the gestures elicited in the previous study.

Task

We presented short video clips (about 2s in duration) of an actor making hand gestures that cast shadows on a static image of a map projected onto a wall. For each video of a shadow gesture, we asked the participant to describe what he thought the actor was trying to do with the map. In the video, the map did not actually respond to the shadow gestures (rendering this task harder than understanding interaction with a working system that provided immediate feedback). Rather than give users options of different effects to choose from, we left it as an open-ended response. This more closely reflected a real-world situation whereby one user has to determine what another user is trying to do based on their shadow.

Participants

We recruited 16 volunteers (8 female) aged 20–57 (median 28.5) from within our institution. Eleven were smartphone owners, and 5 were owners of multitouch mobile phones. No participants were reused from User Study 1, since prior experience could bias participants' understanding.

Gestures

Based on the results of our previous study, we selected a set of gestures for each effect to present to users. Participants in Study 1 performed many distinct gestures, and we aggregated them to create canonical gestures, trying to balance ease of recognition for users and for the system. Hence we recorded videos of an actor performing these canonical gestures, rather than using videos of participants' "raw" gestures captured during the previous study.

The set of gestures in the *Collaborator* condition include 4 panning gestures (*Arm Left*, *Arm Right*, *Arm Up*, *Arm Down*), 3 variations on zoom in and zoom out (*Hand Open* and *Hand Close*, *Arms Open* and *Arms Close*, *Arm Toward* and *Arm Away*), and selection (*Point*). The set of gestures in the *Owner* condition include 4 panning gestures (*Hand Left*, *Hand Right*, *Hand Up*, *Hand Down*), 2 variations on zoom in and zoom out (*Hand Open* and *Hand Close*, *Hand Toward* and *Hand Away*), and selection (*Point*) triggered by dwell.

Experimental Design

In our previous study, people used similar gestures for interacting in the *Maps* and *Photos* conditions, so in this study, we considered only one condition. We arbitrarily decided to use *Maps*. We presented the gesture videos using the same two role conditions as in User Study 1: *Collaborator* and *Owner*. Order of presentation was counterbalanced across participants. Within each condition, operations were presented in random order. Each participant observed 20 shadows: 11 shadows in the *Collaborator* condition, and 9 shadows in the *Owner* condition.

Results

Gender had an effect on goodness ratings with females rating (*median* = 6) the shadows significantly higher (U = 9675.50, z = 3.914, p < .001) than males (*median* = 6).

There were no significant differences in goodness ratings for the different zooming techniques.

We aggregated responses across participants, and found that participants rated the *Owner* condition (*median* = 6) significantly higher (U = 10692.50, z = 2.492, p < .05) than the *Collaborator* condition (*median* = 6).

gesture	pan up	pan down	pan left	pan right	zoom in	zoom out	select	other
hand up	15							draw path
hand down		15						move pin down
hand left			15					move pin left
hand right				16				
hand open					14	1		measure distance
hand toward					15		1	
hand away						14	2	
hand close					2	14		
point				1			15	

Table 2: Effects (column) perceived by participants for each of the different gestures (row) used in the *Owner* condition. Shaded boxes indicate effects we expected to see for each particular gesture.

In a post-study questionnaire, we asked each participant which role condition they preferred overall. Five users preferred *Collaborator*, citing the finer granularity (the hand shadow appears smaller, enabling more precise gestures) and the clarity from two handed usage. Eleven users preferred *Owner*, because the gestures were more clearly defined (framed by the projection area) and more similar to familiar touchscreen gestures.

gesture	pan up	pan down	pan left	pan right	zoom in	zoom out	select	other
arm up	15							rotate
arm down		14						rotate, move pin
arm left			16					
arm right				15				move pin
hand open					16			
arms open					15	1		
arms toward					15			go back to prev location
hand closed					1	15		
arms closed					2	14		
arm away					1	14	1	
point				1			15	pan right

Table 3: Effects (column) perceived by participants for each of the different gestures (row) used in the *Collaborator* condition. Shaded boxes indicate effects we expected to see for each particular gesture.

Table 2 and Table 3 tally the effects that our participants associated with different gestures. On the whole, the effects associated with the different hand shadows reflected similar findings to User Study 1. The outliers associated slightly different effects with certain hand shadows due to ambiguities arising from body mechanics. For example, one person thought the *Collaborator Arm Up* and *Collaborator Arm Down* gestures were intended to rotate the map because, in the video, the user's arm moves up or down relative to the shoulder's axis. Similarly, viewers expressed confusion when gestures for performing inverse operations were not symmetrical. For example, in the videos for *Owner Hand Left* and *Owner Hand Right*, the hand is facing different directions, which feels more natural.

Like in User Study 1, there was some confusion regarding gestures for *Zoom In* and *Zoom Out*. We believe that in an interactive system, this ambiguity will resolve itself.

SHADOWPUPPETS PROTOTYPE

We built a functional prototype to enable us to study the social implications of ShadowPuppets. The system consists of a Logitech 1.3 MP webcam attached to a *Microvision SHOWWX* laser pico projector (Figure 1). Due to low refresh rates associated with the video-out functionality of mobile phones, we used a laptop to drive the display of our prototype . This prototype is intended to simulate a future version running on a mobile phone. The camera and projector are 6cm apart, reasonable for a handheld device.

Gesture Recognition

Our ShadowPuppets prototype was written in C++ using OpenCV. At startup, a calibration step is performed to detect the outer edges of the projected image and the grayscale value of the projection surface. Each frame is thresholded to extract the shadow pixels—pixels with grayscale values close to that of the projection surface.

Connected components within the shadow pixels are detected and then filtered, keeping only the blobs with areas above a threshold, to reduce noise. The system detects a pan or zoom gesture when the average second derivative of the shadow pixels' centroid position or total area over a window of time is above a threshold, and it detects a select gesture when these metrics are below a threshold. We use the second derivatives to avoid the need for an explicit clutching mechanism. The system recognizes one gesture at a time, and a timeout is inserted after recognition to avoid falsely interpreting recoil actions as gestures.

Selection (Point)

To detect pointing shadows, we detect fingertips using an approach similar to Manresa et al. [18]; we look for points on the convex hull that are separated by defects, using the point separated by the deepest defects as our estimate of the fingertip.

When the location of this fingertip is stable over a window of time, we trigger a selection event. The projected map indicates selection by displaying a pop-up box with information about the selected point. The pop-up is removed when another form of input is detected.

Panning (Hand Up, Hand Down, Hand Left, Hand Right)

We track the acceleration of the shadow blob's centroid, taking the average over a sliding window (tuned empirically) of video frames to smooth the sensed data. If the average acceleration in the vertical or horizontal direction is above a threshold, then we fire a *Pan* event (*Up*, *Down*, *Left*, or *Right*), choosing the direction with highest average acceleration. The projected map provides visual feedback, panning in the indicated direction. Similar to existing map applications, we use high initial velocity with constant acceleration in the direction opposite of motion.

Zooming (Hand Toward, Hand Away, Pinching)

Our prototype supports two methods of zooming. Users can either move their shadow towards/away from the wall, or they can perform pinching.

To detect moving the shadow towards/away from the wall we track the second derivative of the change in area of the shadow blob. Again, we use the average over a sliding window for smoothing, and if this average is above a threshold we detect a *Zoom* input event.

Pinching was implemented by detecting finger points, similar to selection. When a transition from one fingertip to two fingertips was observed in close proximity, we fired a *Zoom* event. Unfortunately, in our pilot studies we found that the hand often occluded the shadow in the camera, making detection of this technique unreliable. Because of its unreliability, we left it out of our final user study. Increasing the distance between the camera and projector would reduce occlusion but would also increase the size of the device. A potential future approach could combine shadow and hand detection.

Implementation challenges

Pico projector-based shadow gesture recognition faces some general challenges. Shadow detection requires sufficient contrast between the shadow and the projected content, and may suffer if ambient light is bright.

USER STUDY 3: EXPERIENCES WITH THE SHADOWPUPPETS PROTOTYPE

We conducted a laboratory-based pairs study with our ShadowPuppets prototype to gain insight into users' experiences when using shadow gestures for collocated collaboration. We hoped to learn how it feels to perform the gestures with an interactive system, and what social and technical issues arise when two users interact with the system together. Our prototype supports a simple maps application, that responds to ShadowPuppets hand shadow gestures, and we observed pairs of participants using it.

Task

We first briefly demonstrated how to use each supported gesture to interact with the maps application. The prototype supports 7 different types of shadows (for both roles *Collaborator* and *Owner*): *Hand/Arm Right* to *Pan Right*, *Hand/Arm Left* to *Pan Left*, *Hand/Arm Up* to *Pan Up*, *Hand/Arm Down* to *Pan Down*, *Hand/Arm Toward* to *Zoom In*, and *Hand/Arm Away* to *Zoom Out*.

During the study, one participant performed gestures in the *Collaborator* conditions, while the other participant gestured in the *Owner* condition, as described previously. We asked the participants to interact with the map for 10 minutes using ShadowPuppets, then trade roles (*Collaborator / Owner*) and interact with the map for another 10 minutes. In a post-study interview, we asked participants to describe how it felt to perform each class of gesture: *Panning*, *Zooming*, and *Selection*. We also asked the participants to reflect on their preferences, between the *Collaborator* and *Owner* conditions, and their overall experiences using ShadowPuppets as a pair. We asked participants to think aloud during the task.

Participants

We recruited 8 volunteers (3 female), aged 24–30 (median 27) in 4 pairs of acquaintances, from within our institution. Five participants were reused from User Study 2. Seeing shadow gestures on video beforehand was not likely to bias participants' experiences performing them, since gestures would be demonstrated for training purposes.

Results

All participants learned the gestures quickly; they were able to remember them after viewing a single demonstration of each, and described them as "*intuitive*" and "*making sense*". 7 of 8 participants had positive overall impressions, describing shadow gestures as "*natural*", "*cool*", and "*useful*". The 8th participant explained that he does not like any gestural interfaces because they are imprecise.

Participants envisioned using shadow gestures to interact at-a-distance during presentations while teaching, while in a meeting, or while sharing photos or videos with friends. One participant, a projector-phone user, volunteered that she would want to use shadow gestures individually to get the full benefits of a large projected display, avoiding having to interact with it via the phone's screen.

Panning

All participants volunteered that panning felt "*comfortable*," "*natural*," and "*intuitive*." One participant felt that the gestures were too large, and would prefer if a smaller physical movement would results in a larger movement of the map. Six participants wanted to have more control over how much or how fast to pan, suggesting that the distance and speed of the gesture should correspond to the distance and speed of the pan. Four participants suggested implementing a clutching mechanism, e.g. a hand gesture such as "*grabbing*", to control when a gesture should activate an input event. Four participants experimented with moving the projector relative to the shadow to pan.

Zooming

All the participants described zooming as intuitive and feeling good. One participant (who also participated in the inverse gesture study) commented that he had hated the *hand toward / away* gestures in the videos but that doing it was a lot more intuitive, "*like bringing the map closer to you or pushing it away from you*".

Selection

All the participants liked using pointing for selection. Two participants suggested making a "*tapping*" finger motion to activate selection (as suggested in Study 1). Another participant remarked that shadows are especially good for pointing because it's unambiguous what you are pointing at. He explained that "*when you're in a group and just pointing, the perspective is different for different people,* *but shadows make it much clearer since everyone sees the same shadow*".

Collaborator / Owner Conditions

Six participants preferred the *Collaborator* condition because they didn't have to hold the projector steady while gesturing. Four of the participants said that holding the projector and gesturing felt awkward, and one participant noted that the shadow was too large to control since a small movement had a big effect. Two participants preferred the *Owner* condition because they felt that the person holding the projector had priority to make gestures and had more control: "*I don't think you have as much control if you're not holding it*."

Social Aspects of Collaboration with ShadowPuppets

All the participants liked that more than one person could interact with and manipulate the map, and thought shadow gestures would be good for group interaction. All the participants said that while interacting, they primarily focused their attention on the shadows on the projected display, while attuning to their partners in their peripheral vision.

Because the prototype does not support multiple gestures simultaneously, pairs took turns, negotiating control verbally or by body language ("*if they look like they're going to do something*"). When the pair members both tried to gesture simultaneously, several strategies occurred for managing conflict: they verbally negotiated, both participants backed off and then tried again after an interval, or they "*went over the top of each other and the map did its own thing*". One pair of participants experimented with collaboratively coordinating their actions, with one moving the projector relative to the other's hand.

During the interaction task, one participant in the *Owner* condition intentionally thwarted his partner's gestures by pointing the projector away from her, and unintentionally caused the same problem in the *Collaborator* condition by standing in front of the projected display and occluding it. After the task, when asked how they would prevent the other person from doing something, participants said they would make a confounding gesture ("*making the opposite gesture*", "*waving my arms around*", or "*doing the chicken dance*"), move the projector to point in a different direction (as observed), or occlude the projector by covering it with a hand or standing in front of it.

REFLECTIONS ON SHADOW-BASED INPUT

Initial experiences with ShadowPuppets indicate that the system shows promise for collocated collaboration. Both participants in each pair were able to view and interact with the projected application, maintaining awareness of each other's actions while focusing on the shared display. Study 3 highlights the following issues.

Challenge of Designing for Performers and Observers

Supporting both performers and observers of shadow gestures presents a challenge, since they may have conflicting experiences and goals. To facilitate collocated collaboration shadows should be intuitive and comfortable to perform, and also easy to interpret. Yet these goals may be at odds with one another. For example, users in Study 1 found it easier to control their gestures in the *Collaborator* role, while users in Study 2 found shadows cast in the *Owner* role to be clearer. Similarly, in Study 1 users sometimes performed asymmetric gestures to perform inverse operations in a way that felt comfortable, while users in Study 2 found those gestures harder to interpret.

Need for Fine-Grained Interactions

As a first step toward understanding shadow gestures for interaction, we have examined coarse-grained interaction. Although, we explicitly focused on this part of the design space, users wanted more precision. For example, we selected a set of 7 relatively coarse operations to study and implement, and presented them to users in general terms (e.g., "pan left" rather than "pan left such that the marker on the map is centered"). Yet in Study 3, participants wanted their shadows to perform more precise operations (e.g., controlling the range and speed of panning and zooming), and in Study 2 they often interpreted gestures more precisely than we intended.

Most Popular Gestures May Not Be Best

We posit that the most popular gestures are not necessarily the best. For example, the gestures with highest agreement scores in User Study 1 may simply be the most familiar, since ShadowPuppets uses a new form of interaction. People are biased by what is familiar to them and new ways of interacting may not occur to them immediately. For example, many users drew on pinching gestures from iPhone interaction, while only 25% of users suggested motion in the *throw* dimension (toward or away from the projector). Users' prior expertise also led them to rate pinching gestures higher on the goodness and easiness ratings. Yet in User Studies 2 and 3, we found that *Hand Toward / Away* and *Arm Toward / Away* were intuitive to most users. If we had only considered the results of User Study 1, we might not have even explored those gestures at all. As Norman proposed, user-centered design may not always be the right approach [20].

CONCLUSION AND FUTURE WORK

In this paper we presented ShadowPuppets, a system that allows collocated users to provide input to a mobile projector based system by casting hand shadows.

We made four contributions. First we presented the results of a user study examining what types of shadows users expect to cause different effects. Second, we examined what kinds of effects users expect different hand shadows to cause. Third, we presented the design and implementation of the ShadowPuppets prototype, allowing collocated users to interact with a projected display. Finally, we presented the results of a user study of our prototype, suggesting issues that arise with using shadow-based input.

As future work, we plan to examine how to combine some of the coarse-grained interaction techniques described here with other techniques that provide fine-grained control.

ACKNOWLEDGEMENTS

This work has been funded by AT&T Labs, Microsoft Research ER&P, and UC MICRO 07-067.

REFERENCES

1. Beardsley, P., Forlines, C., Raskar, R., and VanBaar, J. Handheld Projectors for Mixing Physical and Digital Textures. In *Proc. CVPR 2005*, 112.
2. Boring, S., Baur, D., Butz, A., Gustafson, S., and Baudisch, P. Touch projector: mobile interaction through video. In *Proc. CHI 2010*, 2287–2296.
3. Cao, X. and Balakrishnan, R., Interacting with Dynamically Defined Information Spaces using a Handheld Projector and a Pen. In *Proc. UIST 2006*, 225–234.
4. Cao, X., Forlines, C., and Balakrishnan, R. Multi-user interaction using handheld projectors. In *Proc. UIST 2007*, 43–52.
5. Cheng, K. and Takatsuka, M. Initial evaluation of a bare-hand interaction technique for large displays using a webcam. In *Proc. EICS 2009*. 291–296.
6. Cowan, L., Weibel, N., Griswold, W. G., Pina, L., Hollan, J. D. Projector Phone Use: Practices and Social Implications. In *Journal of Personal and Ubiquitous Computing, theme issue on Personal Mobile Projection*, 2011.
7. Echtler, F., Huber, M., and Klinker, G. Shadow tracking on multi-touch tables. In *Proc. AVI 2008*, 388–391.
8. Everitt, K. M., Klemmer, S. R., Lee, R., and Landay, J. A. Two worlds apart: bridging the gap between physical and virtual media for distributed design collaboration. In *Proc. CHI 2003*, 553–560.
9. Freeman, D., Benko, H., Morris, M. R., and Wigdor, D. ShadowGuides: visualizations for in-situ learning of multi-touch and whole-hand gestures. In *Proc. ITS 2009*, 165–172.
10. Greaves, A. and Rukzio, E. Evaluation of picture browsing using a projector phone. In *Proc. MobileHCI 2008*, 351–354.
11. Gustafson, S., Bierwirth, D., Baudisch, P., Imaginary Interfaces: Spatial Interaction with Empty Hands and Without Visual Feedback. In *Proc. UIST 2010*.
12. Harrison, C., Tan, D., and Morris, D. Skinput: appropriating the body as an input surface. In *Proc. CHI 2010*, 453–462.
13. Hilliges, O., Izadi, S., Wilson, A. D., Hodges, S., Garcia-Mendoza, A., and Butz, A. Interactions in the air: adding further depth to interactive tabletops. In *Proc. UIST 2009*, 139–148.
14. Horprasert, T., Harwood D., Davis, L. S. A statistical approach for real-time robust background subtraction and shadow detection. In *Proc. ICCV 1999* Workshops.
15. Ishii, H., Kobayashi, M., and Arita, K. Iterative design of seamless collaboration media. *Commun. ACM* 37, 8 (Aug. 1994), 83–97.
16. Kale, A., Kwan, K., Jaynes, C., Epipolar Constrained User Pushbutton Selection in Projected Interfaces, In *Proc. CVPR 2004 Workshops*.
17. Kane, S. K., Avrahami, D., Wobbrock, J. O., Harrison, B., Rea, A. D., Philipose, M., and LaMarca, A. Bonfire: a nomadic system for hybrid laptop-tabletop interaction. In *Proc. UIST 2009,* 129–138.
18. Manresa, C., Varona, J., Mas, R. and Perales, F. (2005) Hand tracking and gesture recognition for human-computer interaction. *Electronic Letters on Computer Vision and Image Analysis, 5* (3), 96–104.
19. Mistry, P., Maes, P. and Chang, L., WUW - Wear Ur World - A Wearable Gestural Interface. In *Proc. CHI 2009*.
20. Norman, D. A., Human-centered Design Considered Harmful. *Interactions* 12, 4 (Jul./Aug. 2005) 14–19.
21. Pinelle, D., Nacenta, M., Gutwin, C., and Stach, T. The effects of co-present embodiments on awareness and collaboration in tabletop groupware. In *Proc. GI 2008*.
22. Pinhanez, C. S. The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces, In *Proc. Ubicomp 2001*, 315–331.
23. Rapp, S., Michelitsch, G., Osen, M., Williams, J., Barbisch, M., Bohan, R., Valsan, Z., and Emele, M. (2004). Spotlight navigation: Interaction with a handheld projection device. In *Proc. Pervasive 2004, Video Paper*.
24. Segen, J. and Kumar, S. Shadow gestures: 3D hand pose estimation using a single camera, In *Proc. CVPR 1999*, 479–485.
25. Shoemaker, G., Tang, A., and Booth, K. S. Shadow reaching: a new perspective on interaction for large displays. In *Proc. UIST 2007,* 53–56.
26. Snibbe, S. S. and Raffle, H. S. Social immersive media: pursuing best practices for multi-user interactive camera/projector exhibits. In *Proc. CHI 2009*, 1447–1456.
27. Song, H., Grossman, T., Fitzmaurice, G., Guimbretiere, F., Khan, A., Attar, R., Kurtenbach, G. PenLight: Combining a Mobile Projector and a Digital Pen for Dynamic Visual Overlay. In *Proc. CHI 2009*. 143–152.
28. Song, H., Guimbretiere, F., Grossman, T., and Fitzmaurice, G. MouseLight: bimanual interactions on digital paper using a pen and a spatially-aware mobile projector. In *Proc. CHI 2010*, 2451–2460.
29. Sugimoto, M., Miyahara, K., Inoue, H., and Tsunesada, Y. Hotaru: Intuitive Manipulation Techniques for Projected Displays of Mobile Devices. Proc. *INTERACT 2005*, 57–68.
30. Tang, A., Neustaedter, C., Greenberg, S.: Videoarms: embodiments in mixed presence groupware. In *Proc. BCS-HCI*, 2006.
31. Tang, J. C. and Minneman, S. VideoWhiteboard: video shadows to support remote collaboration. In *Proc. CHI 1991*, 315–322.
32. Wilson, A. D. Robust computer vision-based detection of pinching for one and two-handed gesture input. In *Proc. UIST 2006*, 255–258.
33. Wobbrock, J. O., Morris, M. R., and Wilson, A. D. User-defined gestures for surface computing. In *Proc. CHI 2009,* 1083–1092.